

# ALE Analytics - A microbial data warehouse and analytical suite

Patrick V. Phaneuf<sup>1</sup>, Dennis Gosting<sup>2</sup>, Bernhard O. Palsson<sup>2,3</sup>, Adam M. Feist<sup>2,3</sup>

pphaneuf@eng.ucsd.edu, afeist@ucsd.edu

<sup>1</sup> Department of Computer Science and Engineering, University of California at San Diego, La Jolla 92093-0404, United States of America,

<sup>2</sup> Department of Bioengineering, University of California San Diego, La Jolla, CA, 92093, United States of America,

<sup>3</sup> Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800, Lyngby, Denmark

systems biology  
research group



## Abstract

Adaptive Laboratory Evolution (ALE) is a powerful experimental tool used by academic and industrial labs for a number of applications. ALE experiments generate a substantial amount of data, coming in the form of sequencing reads, alignment reports, and sample metadata. As ALE experiments scale to include more samples, the task of managing this data comes at higher costs due to the effort necessary to organize and integrate data into a format that describes the evolution process succinctly. In this work, we describe the development, deployment and iteration of an 'ALE Analytics' pipeline and web platform that streamlines the necessary ALE experiment data post-processing, manages experiment data, and produces interactive reports that detail an ALE experiment. Our design has been primarily driven by the need to consolidate large amounts of ALE experimental data in such a way to describe the quality of the sample sequencing, adaptive mutations in evolved strains, the context of mutations via their metadata (i.e., culturing environments, strain properties), and related mutations found in other experiments housed in the database. We have done so by leveraging a full stack of technologies that enable the parsing and databasing of experiment data, the execution of automated analysis on said data and the generation of web accessible reports. Future efforts will take full advantage of this developing platform to enable more depth and breadth of ALE experiment analysis with quicker turnaround.

## ALE Analytics Pipeline

Our ALE pipeline leverages industry established tools to process and refine ALE replicate sequencing data to identify genomic perturbations and integrate them into a database that can be mined for mutational trends.

